B&P File No. 680-148

BERESKIN & PARR

UNITED STATES

Title: Method and Apparatus for Noise Reduction, Particularly in Hearing Aids

Inventor: Robert Brennan

Title: Method and Apparatus for Noise Reduction, Particularly in Hearing Aids

CROSS-REFERENCE TO RELATED APPLICATION

This application claims benefit from United States provisional application serial no. 60/041,991 filed on April 16, 1997.

FIELD OF THE INVENTION

This invention relates to noise reduction in audio or other signals and more particularly relates to noise reduction in digital hearing aids.

BACKGROUND OF THE INVENTION

Under noisy conditions, hearing impaired persons are severely disadvantaged compared to those with normal hearing. As a result of reduced cochlea processing, hearing impaired persons are typically much less able to distinguish between meaningful speech and competing sound sources (i.e., noise). The increased attention necessary for understanding of speech quickly leads to listener fatigue. Unfortunately, conventional hearing aids do little to aid this problem since both speech and noise are boosted by the same amount.

Compression algorithms used in some hearing aids boost low level signals to a greater extent than high level signals. This works well with low noise signals by raising low level speech cues to audibility. At high noise levels, compression performs only modestly since the action of the compressor is unduly influenced by the noise and merely boosts the noise floor. For persons that frequently work in high ambient sound environments, this can lead to unacceptable results.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a two-fold approach to sound quality improvement under high noise situations and its practical

implementation in a hearing aid. The present invention removes noise from the input signal and controls the compression stage with a cleaner signal. The signal for amplification (the upper path) is, optionally, processed with a different noise reduction algorithm. Under certain circumstances, it may be desirable to use the same noise reduced signal for application and compression control in which case the two noise reduction blocks merge. In another instance, it may be desirable to alter or eliminate the noise reduction in the upper path.

Clearly, noise reduction is not suitable for all listening situations. Any situation where a desired signal could be confused with noise is problematic. Typically these situations involve non-speech signals such as music. A remote control or hearing aid control will usually be provided for enabling or disabling noise reduction.

The present invention is based on the realization that, what is required, is a technique for boosting speech or other desired sound source, while not boosting noise, or at least reducing the amount of boost given to noise.

ins A2 In accordance with a first aspect of the present invention, there is provided a method of reducing noise in a signal, the method comprising the steps:

(1) supplying the input signal to an amplification unit;

(2) subjecting the input signal to an auxiliary noise reduction algorithm, to generate an auxiliary signal;

(3) using the auxiliary signal to determine control inputs for the amplification unit; and

(4) controlling the amplification unit with the control signals, to generate an output signal with reduced noise.

ins A3 Preferably, the input signal is subjected to a main noise reduction algorithm, to generate a modified input signal, which is supplied to the amplification unit. The main and auxiliary noise reduction ~~algorithms can be different.~~

ins A4 In accordance with another aspect of the present

invention, there is provided a method of reducing noise in an input, audio signal containing speech, the method comprising:

(1) detecting the presence and absence of speech utterances;

(2) in the absence of speech, determining a noise magnitude spectral estimate;

(3) in the presence of speech comparing the magnitude spectrum of the audio signal to the noise magnitude spectral estimate;

(4) calculating an attenuation function from the magnitude spectrum of the audio signal and the noise magnitude spectral estimate; and

(5) modifying the input signal by the attenuation function, to generate an output signal with reduced noise.

ins A5 Preferably, the attenuation factor is calculated in accordance with the following equation:

$$H(f) = \left[\frac{|X(f)|^2 - \beta\, |\hat{N}(f)|^2}{|X(f)|^2} \right]^{\alpha}$$

where $H(f)$ is the attenuation function, $|X(f)|$ is the magnitude spectrum of the input audio signal; $|\hat{N}(f)|$ is the noise magnitude spectral estimate, β is an oversubtraction factor and α is an attenuation rule, wherein α and β are selected to give a desired attenuation function. The oversubtraction factor β is, preferably, varied as a function of the signal to noise ratio, with β being zero for high and low signal to noise ratios and with β being increased as the signal to noise ratio increases above zero to a maximum value at a predetermined signal to noise ratio and for higher signal to noise ratios β decreases to zero at a second predetermined signal to noise ratio greater than the first predetermined signal to noise ratio.

ins A6 Advantageously, the oversubtraction factor β is divided by a preemphasis function $P(f)$ to give a modified oversubtraction factor $\beta(f)$, the preemphasis function being such as to reduce β at high frequencies, to reduce attenuation at high frequencies.

Preferably, the rate of the attenuation factor is controlled to prevent abrupt and rapid changes in the attenuation factor, and it preferably is calculated in accordance with the following equation where $G_n(f)$ is the smoothed attenuation function at the n'th time frame:

$$G_n(f) = (1-\gamma)H(f) + \gamma G_{n-1}(f)$$

The oversubtraction factor β can be a function of perceptual distortion.

The method can include remotely turning noise suppression on and off. The method can include automatically disabling noise reduction in the presence of very light noise or extremely adverse environments.

ins A7 ~~Another aspect of the present invention provides for a method of determining the presence of speech in an audio signal, the method comprising taking a block of an input audio signal and performing an auto-correlation on that block to form a correlated signal; and checking the correlated signal for the presence of a periodic signal having a pitch corresponding to that for speech.~~

ins A8 ~~In a further aspect the present invention provides an apparatus, for reducing noise in a signal, the apparatus including an input for a signal and an output for a noise reduced signal, the apparatus comprising: (a) an auxiliary noise reduction means connected to the input for generating an auxiliary signal; and (b) an amplification means connected to the input for receiving the original input signal and to the auxiliary noise reduction means, for receiving the auxiliary signal, the amplification means being controlled by the auxiliary signal to generate an output signal with~~ ~~reduced noise~~.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

For a better understanding of the present invention and to show more clearly how it may be carried into effect, reference will now be made, by way of example, to the accompanying drawings in which:

ins A9 Figure 1 is a conceptual blocked diagram for hearing aid ~~noise reduction~~;

Figure 2 shows a detailed blocked diagram for noise reduction in a hearing aid;

Figure 3 shows a modified auto-correlation scheme performed in segments.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring first to Figure 1, there is shown schematically a basic strategy employed by the present invention. An input 10 for a noisy signal is split into two paths 12 and 14. In the upper path 12, the noise reduction is effected as indicated in block 16. In the lower path 14, noise reduction is effected in unit 18. The noise reduction unit 18 provides a cleaner signal that is supplied to compression circuitry 20, and the compression circuitry controls amplification unit 22 amplifying the signal in the upper path to generate an output signal at 24.

ins A10 Here, the position of the noise reduction unit 18 provides a cleaner signal for controlling the compression stage. The noise reduction unit 18 provides a first generating means which generates an auxiliary signal from an auxiliary noise reduction algorithm. The auxiliary algorithm performed by unit 18 may be identical to the one performed by unit 16, except with different parameters. Since the auxiliary noise reduced signal is not heard, unit 18 can reduce noise with increased aggression. This auxiliary signal, in turn, controls the compression circuitry 20, which comprises second generating means for generating a control input for controlling the amplification unit 22.

The noise reduction unit 16 is optional, and can be effected by using a different noise reduction algorithm from that in the noise reduction unit 18. If the same algorithm is used for both noise reduction processes 16 and 18, then the two paths can be merged prior to being split up to go to units 20 and 22. As noted, the noise reduction in the upper path may be altered or eliminated.

ins A11 With reference to Figure 2, this shows a block diagram of a hearing aid with a specific realization of the proposed noise reduction technique. The incoming signal at 10 is first blocked and windowed, as detailed in applicants' simultaneously filed application serial no. _______ which is incorporated herein by reference. The blocked and windowed output provides the input to the frequency transform (all of these steps take place, as indicated, at 32), which preferably here is a Discrete Fourier Transform (DFT), to provide a signal $X(f)$. The present invention is not however restricted to a DFT and other transforms can be used. A known, fast way of implementing a DFT with mild restrictions on the transform size is the Fast Fourier Transform (FFT). The input 10 is also connected to a speech detector 34 which works in parallel to isolate the pauses in the incoming speech. For simplicity, reference is made here to "speech", but it will be understood that this encompasses any desired audio signal, including music. These pauses provide opportunities to update the noise spectral estimate. This estimate is updated only during speech pauses as a running slow average. When speech is detected, the noise estimate is ~~frozen.~~

As indicated at 38, the outputs from both the unit 32 and the voice detection unit 34 are connected to block 38 which detects the magnitude spectrum of the incoming noise, $|\hat{N}(f)|$. The magnitude spectrum detected by unit 38 is an estimate. The output of unit 32 is also connected to block 36 for detecting the magnitude spectrum of the incoming noisy signal, $|X(f)|$.

ins A12 A noise filter calculation 40 is made based on $|X(f)|$ and $|\hat{N}(f)|$, to calculate an attenuation function $H(f)$. As indicated at 42, this is used to control the original input signal $X(f)$. This signal is subject to an inverse transform and overlap-add resynthesis in known manner, to give ~~an output at 44.~~

During speech utterances, the magnitude spectrum is compared with the noise spectral estimate. In general, frequency dependent

attenuation is calculated as a function of the two input spectra. Frequency regions where the incoming signal is higher than the noise are attenuated less than regions where the incoming signal is comparable or less than the noise. The attenuation function is generally given by

$$H(f) = \left[\frac{|S(f)|^2}{|S(f)|^2 + |N(f)|^2}\right]^{\alpha}$$

where $H(f)$ is the attenuation as a function of frequency
$S(f)$ is the clean speech spectrum
$N(f)$ is the noise spectrum
α is the attenuation rule

The attenuation rule preferably selected is the Wiener attenuation rule which corresponds to α equal to 1. The Wiener rule minimizes the noise power relative to the speech. Other attenuation rules can also be used, for example the spectral subtraction rule having α equal to 0.5.

Since neither $S(f)$ nor $N(f)$ are precisely known and would require a priori knowledge of the clean speech and noise spectra, they are replaced by estimates $\hat{S}(f)$ and $\hat{N}(f)$:

$$|\hat{S}(f)|^2 \approx |X(f)|^2 - |\hat{N}(f)|^2$$

where $X(f)$ is the incoming speech spectrum and $\hat{N}(f)$ is the noise spectrum as estimated during speech pauses. Given perfect estimates of the speech and noise spectra, application of this formula yields the optimum (largest) signal-to-noise-ratio (SNR). Although the SNR would be maximized using this formula, the noise in the resulting speech is still judged as excessive by subjective assessment. An improved implementation of the formula taking into account these perceptual aspects is given by:

$$H(f) = \left[\frac{|X(f)|^2 - \beta |\hat{N}(f)|^2}{|X(f)|^2} \right]^{\alpha}$$

where: β is an oversubtraction factor

α is the attenuation rule

H(f) should be between 0.0 and 1.0 to be meaningful. When negative results are obtained, H(f) is simply set to zero at that frequency. In addition, it is beneficial to increase the minimum value of H(f) somewhat above zero to avoid complete suppression of the noise. While counter-intuitive, this reduces the musical noise artifact (discussed later) to some extent. The parameter α governs the attenuation rule for increasing noise levels. Generally, the higher α is set, the more the noise is punished as X(f) drops. It was found that the best perceptual results were obtained with α = 1.0. The special case of α = 1.0 and β=1.0 corresponds to power spectrum subtraction yielding the Wiener filter solution as described above.

The parameter β controls the amount of additional noise suppression required; it is ideally a function of the input noise level. Empirically it was noticed that under very light noise (SNR > 40 dB) β should be zero. For lower SNR signals, the noise reduction becomes less reliable and is gradually turned off. An example of this additional noise reduction is:

$\beta=0$	for SNR<0
$\beta=\beta_0 \frac{SNR}{5}$	for 0<SNR<5
$\beta=\beta_0 \left[1 - \frac{(SNR-5)}{35}\right]$	for 5<SNR<40
$\beta=0$	for SNR>40

In this example, β_0 refers to the maximum attenuation, 5.0. In effect, from

SNR = 0, the attenuation β is ramped up uniformly to a maximum, β_0, at SNR = 5, and this is then uniformly ramped down to zero at SNR = 40.

Another aspect of the present invention provides improvements in perceptual quality making β a function of frequency. As an instance of the use of this feature, it was found that to avoid excessive attenuation of high frequency information, it was necessary to apply a preemphasis function, $P(f)$, to the input spectrum $X(f)$, where $P(f)$ is an increasing function of frequency. The effect of this preemphasis function is to artificially raise the input spectrum above the noise floor at high frequencies. The attenuation rule will then leave the higher frequencies relatively intact. This preemphasis is conveniently accomplished by reducing β at high frequencies by the preemphasis factor.

$$\hat{\beta}(f) = \frac{\beta}{P(f)}, \quad \text{where } \hat{\beta} \text{ is } \beta \text{ after preemphasis.}$$

Without further modification, the above formula can yield noise reduced speech with an audible artifact known as musical noise. This occurs, because in order for the noise reduction to be effective in reducing noise, the frequency attenuation function has to be adaptive. The very act of adapting this filter allows isolated frequency regions of low SNR to flicker in and out of audibility leading to this musical noise artifact. Various methods are used to reduce this problem. Slowing down the adaptation rate significantly reduces this problem. In this method, a forgetting factor, γ is introduced to slow abrupt gain changes in the attenuation function:

$$G_n(f) = (1-\gamma)H(f) + \gamma G_{n-1}(f)$$

where $G_n(f)$ and $G_{n-1}(f)$ are the smoothed attenuation functions at the n'th and (n-1)'th time frames.

Further improvements in perceptual quality are possible by making β (in addition to being a function of frequency) a function of perceptual distortion. In this method, the smoothing function (instead of a

simple exponential or forgetting factor as above) bases its decision on adapting $G_n(f)$ on whether such a change is masked perceptually. The perceptual adaptation algorithm uses the ideal attenuation function $H(f)$ as a target because it represents the best SNR attainable. The algorithm decides how much $G_n(f)$ can be adjusted while minimizing the perceptual distortion. The decision is based on a number of masking criteria in the output spectrum including:

1. Spread of masking - changes in higher frequency energy are masked by the presence of energy in frequencies in the vicinity - especially lower frequencies;

2. Previous energy - changes in louder frequency components are more audible that changes in weaker frequency components;

3. Threshold of hearing - there is no point in reducing the noise significantly below the threshold of hearing at a particular frequency;

4. Previous attenuation - low levels should not be allowed to jump up rapidly - high levels should not suddenly drop rapidly unless masked by 1), 2) or 3).

For applications where the noise reduction is used to preprocess the input signal before reaching the compression circuitry (schematically shown in Figure 1), the perceptual characteristics of the noise reduced signal are less important. In fact, it may prove advantageous to perform the noise reduction with two different suppression algorithms as mentioned above. The noise reduction 16 would be optimized for perceptual quality while the other noise reduction 18 would be optimized for good compression performance.

A key element to the success of the present noise suppression or reduction system is the speech or voicing detector. It is crucial to obtain accurate estimates of the noise spectrum. If the noise spectral estimate is updated during periods of speech activity, the noise spectrum will be contaminated with speech resulting in speech cancellation. Speech detection is very difficult, especially under heavy noise situations.

Although, a three-way distinction between voiced speech, unvoiced speech (consonants) and noise is possible under light noise conditions, it was found that the only reliable distinction available in heavy noise was between voiced speech and noise. Given the slow averaging of the noise spectrum, the addition of low-energy consonants is insignificant.

Thus, another aspect of the present invention uses an auto-correlation function to detect speech, as the advantage of this function is the relative ease with which a periodic signal is detected. As will be appreciated by those skilled in the art, an inherent property of the auto-correlation function of a periodic signal is that it shows a peak at the time lag corresponding to the repetition period (see Rabiner, L.R., and Schafer, R.W., *Digital Processing of Speech Signals*, (Prentice Hall Inc., 1978) which is incorporated herein by reference). Since voiced speech is nearly periodic in time at the rate of its pitch period, a voicing detector based on the auto-correlation function was developed. Given a sufficiently long auto-correlation, the uncorrelated noise tends to cancel out as successive pitch periods are averaged together.

A strict short-time auto-correlation requires that the signal first be blocked to limit the time extent (samples outside the block are set to zero). This operation is followed by an auto-correlation on the block. The disadvantage of this approach is that the auto-correlation function includes fewer samples as the time lag increases. Since the pitch lag (typically between 40 and 240 samples (equivalent to 2.5 to 15 milliseconds) is a significant portion of the auto-correlation frame (typically 512 samples or 32 milliseconds), a modified version of the auto-correlation function avoiding this problem was calculated. This modified version of the auto-correlation function is described in Rabiner, L.R., and Schafer, R.W., *Digital Processing of Speech Signals, supra*. In this method, the signal is blocked and correlated with a delayed block (of the same length) of the signal. Since the samples in the delayed block include samples not present in the first block, this function is not a strict auto-correlation but shows periodicities better.

It is realized that a hearing aid is a real-time system and

that all computational elements for each speech block are to be completed before the next arrives. The calculation time of a long auto-correlation, which is required only every few speech blocks, would certainly bring the system to a halt every time it must be calculated. It is therefore recognized that the auto-correlation should be segmented into a number of shorter sections which can be calculated for each block and stored in a partial correlation table. The complete auto-correlation is determined by stacking these partial correlations on top of each other and adding as shown in Figure 3.

Referring to Figure 3, input sample 50 is divided into separate blocks stored in memory buffers as indicated at 52. The correlation buffers 52 are connected to a block correlation unit 54, where the auto-correlation is performed. Partial cross-correlations 56 are summed to give the final correlation 58.

This technique quickly yields the exact modified auto-correlation and is the preferred embodiment when sufficient memory is available to store the partial correlations.

When memory space considerations rule out the above technique, a form of exponential averaging may be used to reduce the number of correlation buffers to a single buffer. In this technique, successive partial correlations are summed to the scaled down previous contents of the correlation buffer. This simplification significantly reduces the memory but implicitly applies an exponential window to the input sequence. The windowing action, unfortunately, reduces time periodicities. The effect is to spread the autocorrelation peak to a number of adjacent time lags in either direction. This peak smearing reduces the accuracy of the voicing detection somewhat.

In the implementations using an FFT transform block, these partial correlations (for either technique given above) can be performed quickly in the frequency domain. For each block, the correlation operation is reduced to a sequence of complex multiplications on the transformed time sequences. The resulting frequency domain sequences

can be added directly together and transformed back to the time domain to provide the complete long auto-correlation. In an alternate embodiment, the frequency domain correlation results are never inverted back to the time domain. In this realization, the pitch frequency is determined directly in the frequency domain.

Since the auto-correlation frame is long compared to the (shorter) speech frame, the voicing detection is delayed compared to the current frame. This compensation for this delay is accomplished in the noise spectrum update block.

An inter-frame constraint was placed on frames considered as potential candidates for speech pauses to further reduce false detection of noise frames. The spectral distance between the proposed frame and the previous estimates of the noise spectrum are compared. Large values reduce the likelihood that the frame is truly a pause. The voicing detector takes this information, the presence or absence of an auto-correlation peak, the frame energy, and a running average of the noise as inputs.